How does the methylome relate to clinical outcome?

Methylation arrays add another dimension to the study of COVID-19 and other infectious diseases.

DR. KATHLEEN C. BARNES

PROFESSOR OF MEDICINE & EPIDEMIOLOGY
FOUNDING DIRECTOR, COLORADO CENTER FOR PERSONALIZED MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

SENIOR VICE PRESIDENT, POPULATION HEALTH GENOMICS
TEMPUS LABS

Dr. Kathleen Barnes describes herself as a genetic epidemiologist, studying "how genetic factors can influence disease in populations, typically in the context of the environment, as well as factors influencing its global distribution." This work integrates population with statistical genetics. "My entire research program has almost exclusively focused on disentangling the genetic underpinnings that contribute to health disparities," states Dr. Barnes.

With a rich background in immunogenetics, Dr. Barnes was the ideal candidate to develop and lead the Colorado Center for Personalized Medicine (CCPM) at the University of Colorado (UC) Anschutz Medical Campus. "I spent 23 years as a researcher at Johns Hopkins focused on classic genetic epidemiology and now I'm running a personalized medicine program. I think genetic epidemiology is highly complementary to precision medicine and may have actually led us to where we are with personalized medicine," she states.

The growth of personalized medicine is driven by powerful health insights garnered from multiomics data. As a resource for collecting this data, CCPM manages an extensive biobank comprising nearly 200,000 patient samples. When the COVID-19 pandemic arose in March 2020, Dr. Barnes and the CCPM were well positioned with patient samples to start projects related to coronavirus infection. "On day two of the shutdown, the CCPM biobank was asked to pivot and commence with COVID testing." Dr. Barnes and her team were also able to continue with their research activities and uncovered connections between host methylation, SARS-CoV-2 infection, and clinical outcomes.

We had the opportunity to meet with Dr. Barnes and discuss her work at the CCPM and how the methylome may contribute to infectious disease.

For Research Use Only. Not for use in diagnostic procedures.

M-GL-00158 v1.0 | 1

The many "firsts" achieved by the CCPM:

- FIRST institutional biobank to leverage genomics data from a commercial SNP array
- FIRST program to launch clinical integration of pharmacogenomics through a research biobank
- FIRST enterprise health data warehouse to partner with Google Cloud platform

Q: What are the goals of the Colorado Center for Personalized Medicine?

Kathleen Barnes (KB): The goal of CCPM is to personalize care for every individual or patient, typically leveraging genetic information within the CCPM. The goal is linking extensive electronic medical record data to omics information. We've recruited and built up a base of faculty that have expertise in the field of genetic epidemiology to help achieve these goals.

Another goal of CCPM is to establish strategic partnerships that cover end-to-end deployment of clinical applications towards individualized health. This includes, but it's not limited to, reliable, feasible, scalable testing assays, diagnostic predictive algorithms, and clinical decision support pipelines, all of which really enhance our existing capabilities and enable our program to reach economies of scale.

Q: What is the CCPM Biobank?

KB: This is our CLIA- and CAP-certified biobank done in collaboration with our partner/healthcare system, UCHealth.* We've consented nearly 200,000 UCHealth patients to the biobank since we launched in the spring of 2016. The recruitment process is all done electronically through the UCHealth patient portal where patients have the opportunity to consent. It may be some time before they come back in and we have the opportunity to collect a blood sample. To date, we've either already extracted, or we're currently extracting samples from, about 80,000 participants.

Q: Before COVID, what was the main research focus of your lab?

KB: My entire research program has truly focused, almost exclusively, on disentangling the genetic underpinnings that contribute to health disparities. My main focus has been in complex lung diseases and allergic diseases and other diseases of inflammation—most notably asthma, but almost exclusively focused on populations of African ancestry.

This project is NIH supported through a grant called Consortium on Asthma Among African-ancestry Populations in the Americas (CAAPA). In 2011, we shifted from exome sequencing to whole-genome sequencing at a reasonably affordable cost. We

_____

* CLIA, Clinical Laboratory Improvement Amendments; CAP, College of American Pathologists; UCHealth University of Colorado Hospital

For Research Use Only. Not for use in diagnostic procedures.

were able to sequence whole genomes for about 1000 individuals of African ancestry, representing North/Central/South America, the Caribbean, and Africa. The goal was to generate the African Diaspora Genomic Catalog. We used that catalog to partner with Illumina and develop what we called the African Power Chip. The idea was to create a customized SNP[†] chip that would allow us to fill in the gaps for information that wasn't captured on a commercial chip.

Illumina then decided to create the Multi-Ethnic Genotyping Array, or the MEGA chip. We partnered closely with Illumina, as well as another NIH consortium called PAGE [Population Architecture using Genomics and Epidemiology], to take all the content we had created in the African Power Chip and include it on the MEGA Chip, as well as clinically actionable sorts of data. We hadn't imagined this would happen and it was very exciting. That has been one of the most gratifying accomplishments, of my career, but also on behalf of CAAPA. We're very proud of the CAAPA catalog. To this day, it serves as an invaluable resource for the research community at large because it enables imputation of genetic variation in African ancestry populations at very high accuracy.

We're also gratified that we were able to develop a GWAS[‡] chip that captures ancestral diversity, not just of the African genome, but of multiple non-European populations. This continues to shape my research as we move into a phase of multiomics where we're fortunate to have this GWAS-level data, and even whole-genome sequence data, as the backbone. From that backbone, we can then layer on transcriptomics, methylomics, proteomics, metabolomics, etc. It's been very exciting.

Q: How did COVID impact your research plans?

KB: It was challenging when COVID happened in March 2020. I think it especially impacted programs like CAAPA because they require recruitment of patients with and without a particular disease, and there was an absolute halt on any kind of clinical research across all the sites, including our international partners in Barbados, Brazil, and Nigeria. The CCPM Biobank was asked to pivot and leverage our infrastructure to commence with COVID testing. Having a fully running biobank meant that we were able to continue with most of our CCPM activities during that period.

"This continues to shape my research as we move into a phase of multiomics where we're fortunate to have this GWAS-level data, and even whole-genome sequence data, as the backbone. From that backbone, we can then layer on transcriptomics, methylomics, proteomics, metabolomics, etc."

---

† SNP, single nucleotide polymorphism

‡ GWAS, genome-wide association study

In the research lab, which we call CARGO, we were allowed to function at full capacity because of our focus on developing a novel, methylation-based COVID-19 diagnostic platform.

Q: What sparked your interest in looking at methylation data in the context of infectious disease?

KB: In the months leading up to the pandemic, our team at CCPM, along with our Illumina partners, was considering how we might introduce clinical methylation-based testing into the CCPM Biobank, specifically for tumor classification and diagnostics. This was really inspired by the interesting work that had been published around developing classifiers and diagnostics of neuro tumors and other types of cancers. This was all done on the Illumina Infinium™ MethylationEPIC BeadChip.

Immediately after the shutdown, there was a GenomeWeb story that came out about studying the epigenetics of COVID. The consideration of generating disease classifiers using this methylation platform was fresh in my mind so I immediately called my closest Illumina counterpart and I said, "Alem [Taye], we could do this." It seemed reasonable, if we could leverage the biospecimens. That was one of the biggest challenges in the early months of the pandemic at every institution. There was a mad scramble to access biospecimens from COVID patients for research. There was sort of a feeding frenzy around these samples and, to be honest, that was a big challenge for us in thinking about how we could stand this up.

Initially, we focused on collecting leftover samples from nasal pharyngeal samples collected for COVID PCR testing. But we theorized, because it's infectious disease, we should be able to do this in blood. Surprisingly, collecting blood turned out to be much easier than trying to capture these leftover samples. We also decided that the gold standard for confirmation of disease status of these patients would be conventional PCR. Through our health data warehouse, we were able to track and confirm whether samples that came into our research lab had a paired PCR test, and if it was positive or negative. We also customized the MethylationEPIC platform by adding another 7800 CpGs representing genes that we thought would be relevant, not just for SARS-CoV-2 infection, but for infectious disease in general, and we started calling this the MethylationEPIC+ BeadChip.

"We also customized the MethylationEPIC platform by adding another 7800 CpGs representing genes that we thought would be relevant, not just for SARS-CoV-2 infection, but for infectious disease in general, and we started calling this the MethylationEPIC+ BeadChip."

Q: Why would DNA methylation changes be associated with COVID-19? And what does this mean for the host?

KB: There are studies out there that demonstrated that enveloped RNA viruses, such as SARS-CoV-2, can manipulate the host epigenome, specifically via DNA methylation. There was a lot of evidence at the time, and more so now, that you can develop highly accurate and robust machine learning–based disease classifiers using DNA methylation patterns. While there's been a lot of work to study infectious disease at the level of gene expression, for example, transcriptomics and germline susceptibility at the level of the genome, we really felt like the methylome hadn't been fully exploited to the extent that it could be in this space. I think an advantage of DNA methylation profiling over RNA transcriptome profiling in developing classifiers for disease state, is in fact the stability of the deoxyribonucleic acid over ribonucleic acid, and just the inherent instability of RNA. In addition to that, methylation quantification, using a tool like the EPIC chip, is actually pretty affordable at scale. Being able to study the relationship between methylation patterns, in the presence or absence of a disease, is potentially as informative as the other types of diagnostics platforms and we always imagine that this could potentially be scaled in whole communities to really address the pandemic.

What's really unique about this platform, and really inspired us to move forward with the study, is if you think about the other diagnostic tests that were emerging rapidly at the time of the onset of the pandemic, sequencing, RT-PCR, or antibody tests, the methylation platform can also create these disease classifiers that can predict outcome of disease, not just diagnosis.

One of our major findings is that DNA methylation profiling in conjunction with analysis using machine learning techniques can, in fact, identify a SARS-CoV-2 specific epigenetic signature. We've demonstrated this in blood. We also describe the development of a classification algorithm and we've been able to demonstrate high sensitivity and specificity in predicting an infection, and also in-hospital clinical deterioration. Our findings also suggest that the measurement of methylation signatures that come up during and after SARS-CoV-2 infection can provide clinicians with the ability to detect viral infection, as well as predict patient clinical course, which is a real differentiator that would potentially enable clinicians downstream to triage their patients with clinical decision support.
.

"...the methylation platform can also create these disease classifiers that can predict outcome of disease, not just diagnosis. For us, being able to predict a patient's clinical course after a viral challenge is a real differentiator, and that would potentially enable clinicians downstream to triage their patients with clinical decision support."

For Research Use Only. Not for use in diagnostic procedures.

M-GL-00158 v1.0 | 5

> "we used EWAS results to validate our findings from the process of these disease classifiers...even though this proof of concept was limited to a single RNA enveloped infectious agent, that is SARS-CoV-2, it is certainly reasonable that this technology could be leveraged across other respiratory infectious diseases."

Q: Why did you choose the Illumina MethylationEPIC BeadChip?

KB: I've been an Illumina user for a very, very long time, going back to the GoldenGate™ days. We embraced the Infinium chemistry platform when it first came out. Initially focused on the GWAS chips, but, of course, it's interchangeable for the EPIC chip. In our research space, we've been working with the EPIC chip§ for several years. We had the infrastructure in place and the know-how and the expertise to work with this platform. We're very familiar with how the platform has been developed and the selection of CpGs, and we feel very comfortable with that. And absolutely, with the ongoing relationship with Illumina, it was a very easy pivot for us.

Q: How are you analyzing the data?

KB: Machine learning is something near and dear to our hearts in the clinical informatics space. We have quite a bit of expertise in the center with faculty who routinely leverage machine learning, again around clinical informatics primarily. We use algorithms to parse data and we learn from it and make a determination, or we make a prediction about the future state of any new data sets. So, instead of the old-fashioned way of doing things, with a programmer hand-coding software with a specific set of instructions, in machine learning the machines train using large amounts of data and algorithms and that then gives it the ability to learn how to perform the task. The programmer codes the algorithm used to train the network instead of coding expert rules. And then the algorithms, of course, improve their performance as the quantity and the quality of the data are available and learning increases. It builds a model from sample inputs and it uses that model to make predictions based on subsequent data. We very much looked to the success of leveraging the EPIC chip in the cancer world for creating disease classifiers to differentiate tumors.

Q: How do you see the different omics being combined as we move forward with infectious disease research?

KB: In thinking about the work that we've done with COVID, certainly combining the methylation array with GWAS arrays has proven to be a cost-efficient method. Combining the genomic data with any of the other omics, through QTL* mapping, is a powerful

---

§ Including previous versions of the MethylationEPIC BeadChip, eg, the Infinium 450K and 850K BeadChips

¶ QTL, quantitative trait locus

approach. In fact, in our paper[1], we used EWAS** results to validate our findings from the process of these disease classifiers. You're looking for an epigenetic signature. And I think, even though this proof of concept was limited to a single RNA enveloped infectious agent, that is SARS-CoV-2, it is certainly reasonable that this technology could be leveraged across other respiratory infectious diseases.

Q: What are the next steps for you?

KB: We want to expand the COVID methylation project to develop disease classifiers that can identify asymptomatic individuals. And we're really focusing on our capabilities around differentiating SARS-CoV-2 from other respiratory viruses. We'll continue to build our biospecimen repository. As this data set grows, there will be a need to expand our partnerships. There's a limit to what an academic center can accomplish around really robust machine learning, and artificial intelligence in general. We recognize the need for industry partnerships and working with companies that have built robust machine learning platforms already.

In terms of the next step for the center, we'd been having discussions about how we could implement MethylationEPIC into our clinical biobank in the space of oncology, but now we're thinking much bigger. Ideally, we will use this EPIC+ chip across our entire CCPM biobank population, really focusing on infectious disease. We think this is something feasible and novel, and we have existing GWAS data on the population already.

"Ideally, we will use this EPIC+ chip across our entire CCPM biobank population, really focusing on infectious disease. We think this is something feasible and novel, and we have existing GWAS data on the population already."

Reference

1. Konigsberg IR, Barnes B, Campbell M, et al. Host methylation predicts SARS-CoV-2 infection and clinical outcome. Preprint at www.researchsquare.com/article/rs-334297/v1. Posted March 22, 2021. Accessed September 8, 2021.

** EWAS, epigenome-wide association study

Read the paper

Konigsberg IR, Barnes B, Campbell M, et al. Host methylation predicts SARS-CoV-2 infection and clinical outcome. Preprint at www.researchsquare.com/article/rs-334297/v1. Posted March 22, 2021. Accessed September 8, 2021.

Learn more

Infinium MethylationEPIC BeadChip, www.illumina.com/products/by-type/microarray-kits/infinium-methylation-epic.html

Methylation changes and COVID, www.illumina.com/content/dam/illumina-marketing/documents/icommunity/methylation-covid-barnes-case-study-m-gl-00167.pdf

COVID-19 host risk and response, www.illumina.com/areas-of-interest/microbiology/infectious-diseases/host-genetics-immune-response.html

illumına®